Data Review Checklist

The release of USGS data and data products must be as universal as scholarly publications to comply with recent mandates to make scientific data accessible. To achieve this goal with consistent quality standards, the data review must be included in the review and approval process prior to release.

Data review is similar to peer review, but different in some important ways. Both evaluate a product intended for public release; in this case the product is data rather than a scholarly publication. Both reviewers perform a detailed evaluation of the product, provide comments and suggestions to the author, and certify that the review has occurred. The special focus of the data reviewer is the accuracy, completeness, and usability of a data product. The basic questions to be answered by the data review are: Are the data what the author says they are? Do they meet specifications for quality, accuracy, and completeness as identified by both the author and the approving official?

The following checklist is provided for the assistance of data reviewers who are experienced in working with the particular type of data being reviewed. It is assumed that data reviewers have the expertise and tools to access and assess both the data and the metadata, and also are knowledgeable about appropriate methods for managing and presenting data within the relevant scientific discipline.

In some cases, it will be unreasonable to actually check every data value, so a spot check or a check of a carefully selected sample may need to suffice. In every case, the data reviewer should include in their report a description of how much of the data, and which data, they examined. If the product requires a peer review as well as a data review, the same reviewer can perform both. The data reviewer may also serve as the metadata reviewer (http://www.usgs.gov/datamanagement/share/datarelease.php).

A data review may include checks for the following:

- Are data values reasonable?
  - Are they in a valid range for that measurement?
  - Do they display seasonal or daily trends that are applicable?
  - Is there consistency between adjacent or otherwise related datasets, within the product?
  - Are the geographic locations given for the data, reasonable? (i.e., are ocean data points actually showing in the ocean?)
  - Is the accuracy claimed for the data reasonable?
  - Are “no data” values accurately defined?
  - Are data anomalies or gaps explained in the metadata?
  - Do analysis values add up?

- Are the data development methods scientifically sound and well described?
  - Could a scientist or technician recreate the final data set from the descriptions?
  - Can the documentation about methodology be easily found and used?
  - Are processing software and versions identified?
- Does the product include metadata in a standard endorsed by the Federal Geographic Data Committee (FGDC) for all the data, such as the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) or ISO 19115?
- Are the coordinate system and datum defined appropriately (both horizontal and vertical)?
- Does the Product as a whole, through its design or documentation, provide enough information so that the data and metadata can be easily found and used?
  - Is the format of the file identified?
  - Is there information about the software required to use the data?
  - Is the location of files documented?
  - If the data are released on a web page, does the page have useful discovery metadata, for example the web page clearly identifies the contents, keywords and metadata tags are provided, and geospatial attributes are presented?